A Bioinformatics Analysis of the Relationship between Stowaway-like Transposable Elements and Scaffold/Matrix Attachment Regions

An Honors Thesis (HONRS 499)

by

Anna E Grybas

Thesis Advisor

Dr. C Ann Blakey

Ball State University

Muncie, Indiana

05/2011

Expected Date of Graduation

05/07/2011

ABSTRACT

Predictions of a relationship between transposons and S/MARs were made by Tikhonov, Bennetzen, and Avramova (2000) and Christoffers (1998). This analysis compares the presence of transposons from the Stowaway family relative to two specific plant S/MARs, pTMX4 and pSR14, isolated by Christoffers (1998). Homologous sequences were identified using both pTMX4 and pSR14 by BLASTn and MEGABlast searches. These sequences were further analyzed for the presence of Stowaway elements. General Stowaway patterns were observed with respect to the level of homology to the S/MAR, the species, and the quantity of Stowaways per sequence. These patterns suggest there is a positive relationship between Stowaway quantity and the level of S/MAR homology. Detailed Stowaway patterns were observed among individual Stowaways within the sequences with respect to location, order, and fragmentation of individual Stowaway elements. These Stowaways revealed probable insertion patterns of nested Stowaway elements and models were subsequently generated to represent them. The complete description of the Stowaway analysis of pTMX4 and SR14 are provided.

ACKNOWLEDGEMENTS

First and foremost, I'd like to thank my parents, Loreta Tvrdy and Don Grybas, for being my support system through the past four years I have spent here, at Ball State University. It is with their support that I have been able to attend and remain composed, even in the hardest times. Because of them, I have felt the confidence to take on academics, including this project.

I would like to thank my research advisor, Dr. Ann Blakey for giving me the opportunity to take on this project and assisting me when I came across problems. I appreciate the help in writing the final draft to this paper, as well.

TABLE OF CONTENTS

LIST OF FIGURES

DETAILED LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

Background

Not much is known about the presence of Stowaway-like transposable element patterns, particularly with regard to scaffold/matrix attachment regions (S/MARs) and the surrounding sequences. By analyzing sequences homologous to S/MARs for the presence of Stowaway-like transposons, the relationship between transposable elements and S/MARs can be better understood. These analyses can be performed using bioinformatics technology which can expose more uses for bioinformatics technology, spread the importance of bioinformatics databases, and determine any areas within the databases that could be improved for more effective use. To address these concerns, I propose a bioinformatics analysis of Stowaway-like transposable elements associated with S/MARs.

S/MARs are sequences of DNA that allow the strand of DNA to fold into loops and attach to a protein skeleton. This is the process DNA goes through to form the chromosome. S/MARs are one element of DNA that allows for chromosomal rearrangements. Transposable elements are sequences of DNA which excise and insert themselves on chromosomes. The triggers causing excision and recognitions sequences for insertion are not fully understood. Some transposable elements can excise and re-insert themselves multiple times, often leaving a trace behind like a sequence fragment or a whole copy of the sequence. Understanding the relationship between S/MARs and transposable elements may help explain certain forms of chromosomal rearrangements and thus gene expression and silencing. To identify the sequences of transposable elements located

near homologous sequences of S/MARs in graminoid genomes, bioinformatics technology can be used.

Bioinformatics technology is the computerization of results and research which allow biological data to be globally accessible. The accessibility is important in research because the ease of retrieving results helps research advance quickly without repeating studies and allowing the globe to work on large projects together to benefit the world. Computerization of data allows for quick comparisons of new data as well. Bioinformatics is an extremely valuable field with an application in nearly all forms of science. Currently, researchers are using bioinformatics before research begins to gather the appropriate information to start a project. Bioinformatics technology is used to compare collected data, publish and present research globally. Since bioinformatics technology is still in its youth, there is a lack of standard data entry, which can make comparing and analyzing data difficult. Though all the data is still available, a standard entry method would allow for even speedier comparisons.

Scaffold/Matrix Attachment Regions

Scaffold/Matrix attachment regions (S/MARs) are sequences of DNA that bind to a protein skeleton during chromosome formation, causing the strand of DNA to create loops. S/MARs are unique to chromosomal DNA and are not contained on organelle DNA, even after the DNA has migrated into the nucleus (Rudd et al 2004). According to Heng et al (2004), the loops formed by the S/MARs are dynamic, controlled by an unknown regulatory system, which selects which S/MARs are used to form loops. The fluidity of these loops appears to be associated with the function of S/MARs in gene expression and chromatin formation. S/MARs identified to occur only once in the genome were found to always bind, the S/MARs in multiple copies participate in the dynamism of the loop formation (Heng et al 2004).

The effect of S/MARs on the genome is not completely understood, however S/MARs appear to be involved in gene regulation and chromosomal rearrangements. In the model created by Heng et al (2004), they explain that the S/MAR encourages an association between the nuclear matrix and the gene, which is necessary during gene expression or replication. Additional studies suggest that S/MARs reduce gene silencing (Allen, Spiker, and Thompson 2000; Brouwer, Bruce, Maddock, Avramova, and Bowen 2002). And there is evidence which shows when a gene is flanked by a S/MAR, the expression of that gene may increase (Brouwer, Bruce, Maddock, Avramova, and Bowen 2002).

In general, S/MARs are identified and classified based on binding strength (Gasser and Laemmli 1986; Tikhonov, Bennetzen, and Avramova 2000). Binding strength can be measured using *in vitro* binding assays (Brouwer, Bruce, Maddock, Avramova, and Bowen 2002). S/MAR sequences appear to be A-T rich and associated with histone proteins (Liebich, Bode, Reuter, and Wingender 2002; Christoffers 1998; Gasser and Laemmli 1986). A database was created by Liebich, Bode, Frisch, and Wingender (2001) to organize and categorize proteins identified to associate with S/MAR sequences in order to identify relationships between S/MARs, proteins, and gene expression. Unfortunately, this database is no longer active.

Research by Tikhonov, Bennetzen, and Avramova (2000) found some S/MARs in maize are located between genes (*adh*) and highly repetitive DNA. It appears that the S/MAR acts as a spacer to separate the two. This study also identified S/MARS which flank genes. S/MARs are not typically found evenly distributed across the chromosome, thus creating uneven loops (Linnemann, Platts, and Krawetz 2009). Small loop forming S/MARs do not appear to have a preference of placement in the genome with regard to genes or gene density (Linnemann, Platts, and Krawetz 2009). S/MARs were identified in *Arabidopsis thaliana* at a density of approximately one S/MAR contained within every 5.5 kb of DNA (Rudd et al 2004).

The S/MARs used for analysis in this research were isolate by Christoffers (1998). Specifically, this work focused on two wheat S/MARs: 1) pSR14 [GenBank: AF176229.1], a subclone of λAmy3/33 originally isolated by Baulcombe et al (1987) and subcloned by Christoffers (1998); 2) pTMX4 isolated by Christoffers (1998). The S/MAR from the clone pSR14 contains an α-amylase gene, *Amy3,* and the corresponding 5' flanking region. The pSR14 sequence was identified as AT-rich and found within a microsatellite. The second S/MAR, pTMX4, was also AT-rich and found to have homologous sequences in Barley and Rye (Christoffers 1998). Understanding the relationship between pSR14, pTMX4, and Stowaways may help support and explain their functions. Evidence from previous studies supported a predicted relationship between transposons and S/MARs (Tikhonov, Bennetzen, and Avramova 2000; Christoffers 1998), in particular Christoffers (1998) noted a potential relationship between Stowaway-like transposable elements and S/MARs.

Transposable Elements

Like scaffold/matrix attachment regions, transposable elements, commonly called transposons, are another unique feature of genomes. These elements compose the largest portion of many different eukaryotes and over 80% in some plants (Feschotte, Osterlund, Peeler, and Wessler 2005; Durand and Michod 2010). Found in nearly every genome, transposable elements are segments of DNA which are capable of insertion, excision, and re-insertion into genomes. These transposition activities can occur within the same genome or between genomes (Lewin 2008a). As transposable elements transpose within a genome, the sequences can leave fragments behind or bring pieces of the original genome to the next genomic region. This can lead to genome rearrangements which may have a phenotypic effect on the organism it resides within (Brown 2007, Lewin 2008a). Transposons can move through the genome using transposases, enzymes which assist in the movement of the element within or between genomes.

According to Feschotte, Osterlund, Peeler, and Wessler (2005), several different categories of transposable elements exist. One division between transposons is class; class 1 and class 2. Class 1 transposons are called retrotransposons which utilize an RNA intermediate and reverse transcriptase to transpose. Class 2 transposons transpose DNA to DNA and utilize transposases (Feschotte, Osterlund, Peeler, and Wessler 2005). Another description for transposable elements describes the size and placement of the element. For example, miniature inverted-repeat transposable elements, or MITEs, are most commonly found in high quantities in plant genomes (Macas, Koblizkova, Neumann 2005). In animals, there are often long interspersed repeated sequences (LINES) and short interspersed repeat sequences (SINES) (Lewin 2008b). Within the MITE category, the transposable elements are divided into families. One family within the MITEs is the Stowaway family.

Stowaway Family of Transposable Elements

Stowaway transposable elements are found in many plant genomes ranging through all cereal grasses (Bureau and Wessler 1994). A complete report of all Stowaways located in a genome is available for very few plant genomes (Macas, Koblizkova and Neumann 2005). These genomes including rice (Feschotte, Osterlund, Peeler, and Wessler 2005) and pea (Bureau and Wessler 1994). Maize, barley, and wheat genomes have also been studied for Stowaway content (Feschotte, Swamy, and Wessler 2007). Macas, Koblizkova and Neumann (2005) concluded that due to an association of Stowaway elements and genes, Stowaways may make good biomarkers one day. Rice is a common plant studied for MITEs, like Stowaway transposable elements, and the transposases associated with them (Feschotte, Osterlund, Peeler, and Wessler 2005). Stowaways are one of the most common types of transposons found in rice (Bureau, Ronald, and Wessler 1996). The vast amount of information available in the rice genome makes this genome popular to study.

Stowaways were initially discovered, by Bureau and Wessler (1994), due to the structural similarity to the Tourist family of transposable elements. Stowaways and Tourists are both found in the cereal grasses, however, Tourists are limited to only certain cereal grasses whereas Stowaways have been identified in the two major plant classifications, monocot and dicot. Further investigation revealed that there were no sequence similarities between Tourist and Stowaways. During the beginning research of Stowaway transposable elements, 47 sequences were identified to contain a Stowaway. Fifty different Stowaways were identified after the discovery of Stowaways based on at least 60% sequence similarity, similar terminal inverted repeat sequences, overall size, length, and structure (Bureau and Wessler 1994). One method of identifying other Stowaway sequences is to use the sequence of a previously identified Stowaway transposable element and create a primer. This primer can be used to identify homologous sequences which can be further investigated to identify Stowaways (Macas, Koblizkova, and Neumann 2005).

Stowaways are a type of MITE, miniature inverted-repeat transposable element. MITEs are non-coding pieces of DNA found in various non-coding regions of the genome. These sequences form hairpin structures in the genome (Bureau and Wessler 1994). Stowaway transposable elements are typically characterized by an 11 bp terminal inverted repeat containing the sequence 5′ – CTCCCTCCGTT – 3′ (TIRs) and 5′ – TA – 3′ target site duplications (TSDs) (Feschotte, Osterlund, Peeler, and Wessler 2005; Bureau, Ronald, and Wessler 1996; Bureau and Wessler 1994). These characteristics of the TIRs and TSDs allowed many plant MITEs to be divided between two distinct groups, Stowaway and Tourist (Feschotte, Osterlund, Peeler, and Wessler 2005). Stowaways range from approximately 80 – 323 bp in length (Bureau and Wessler 1994). Stowaway transposable elements are AT rich (Bureau and Wessler 1994). A study of the pea genome completed by Macas, Koblizkova and Neumann (2005) discovered that there are slightly more than 100 Stowaway transposable elements in each haploid pea genome. Turcotte, Srinivasan, and Bureau (2001) studied

the rice genome and identified 70 Stowaways. Within the Stowaway classification group, these elements have been further divided into 36 different families in the pea genome and 24 groups in the rice genome (Feschotte, Osterlund, Peeler, and Wessler 2005 and Turcotte, Srinivasan, and Bureau 2001). An attempt to classify Stowaways was completed by Macas, Koblizkova, and Neumann (2005) in which the sequences surrounding the Stowaway were analyzed, however this was not as successful as hoped due to a lack of sequence and classification information provided for the surrounding sequences. Inadequate information and inconsistency in annotation are still issues in this field of research.

One common inconsistency in Stowaway transposable elements is naming. Prior to obtaining the Greek names which are often used now to distinguish Stowaways, Stowaways were named with the prefix *Stow* followed by hyphenated initials of the genus and species paired with a number representing the transposable element (Macas, Koblizkova, and Neumann 2005). Stowaways are inconsistently labeled today with some containing purely a "MITE" label, some "Stowaway" and others with the Greek names like Athos, Thalos, and Icarus to name a few.

Stowaway transposable elements are theorized by Feschotte, Osterlund, Peeler, and Wessler (2005) to be transcribed using transposases transcribed from Osmars, mariner-like elements. This theory arose from the identified similarities of the TIRs of both Stowaway and Osmar or Psmar elements (Macas, Koblizkova, and Neumann 2005; Feschotte, Osterlund, Peeler, and Wessler 2005).

A study by Macas, Koblizkova, and Neumann (2005) attempted to characterize Stowaways by identifying them and analyzing the sequences around them for commonalities or characteristics; it has been noted that Stowaway elements are associated with gene sequences. In a reciprocal approach, our study identified the sequence around one of two S/MARs and then analyzed the associated Stowaway transposable elements. Approaching the analysis of Stowaways from both

directions allows for greater understanding of the significance and functionality of Stowaways with particular gene sequences.

Bioinformatics Analysis

The use of bioinformatics is dramatically increasing as science progresses. Bioinformatics allows for the globalization of data and research by utilizing the internet and databases. Some of these databases include GenBank and EMBL (Bureau and Wessler 1994; Bureau, Ronald, and Wessler 1996). Both these databases are utilized by searching a sequence of DNA against the database. The database reveals sequences with a pre-determined level of homology to the starting entry sequence. In GenBank, this search is called a Basic Local Alignment Search Tool or BLAST search. Within the National Center for Biotechnology Information (NCBI) database, several difference BLAST searches can be performed. A BLASTn search performs a comparison of a nucleotide sequence submitted by the user to sequences within the database. A MEGABlast search through NCBI also performs a nucleotide comparison requiring a higher level of homology in order to obtain results. These two BLAST searches were used to complete this analysis.

These searches using these types of tools have allowed for collaboration on sequences which in turn has led to greater annotation and classification of genomic elements, such as S/MARs, MITEs, Stowaways, and many others (Macas, Koblizkova and Neumann 2005).These searches have also provided easier means to compare the genomic elements (Feschotte, Swamy, and Wessler 2007). Some specific databases have been created outside the more commonly used ones, such as the one for S/MARs created by Liebich, Bode, Frisch, and Wingender (2002). Some of these databases are no longer active, such as the S/MAR database previously mentioned, possibly due to incorporation into larger public databases. With increasing ease and accessibility, there has been a proliferation in the ability to characterize elements such as Stowaways within sequences housed

within databases (Feschotte, Swamy, and Wessler 2007). Other than the highly curated data sets, there appears to be minimal standardization of the annotation systems used in the characterization of submitted sequences and their elements. This lack of standardization can lead to a new type of complication or problem associated with the advancement of bioinformatics.

CHAPTER 2

MATERIALS AND METHODS

Materials

The scaffold/matrix attachment region (S/MARs) sequences used for this study were pTMX4 and pSR14, first identified by Christoffers (1998). They were used in bioinformatics searches to identify additional putative S/MAR sequences in wheat and other cereal grasses. The regions identified to contain a putative S/MAR sequence, based on homology to either pTMX4 or pSR14, were then used for further analysis to identify the presence of Stowaway-like transposable element patterns in the surrounding sequences.

The Basic Local Alignment Search Tool (BLAST), found at the National Center for Biotechnology Information (NCBI), was used to perform searches to find the sequences with homology to pTMX4 or pSR14. Two different nucleotide BLASTs were used, BLASTn and MEGABLAST, which complete the same searches with the only difference being the degree of homology required. BLASTn searches for homologous sequences allowing some variability with regard to gaps and mismatches there by only identifying somewhat similar sequences. MEGABLAST searches are preformed to identify sequences that have a high degree of similarity, fewer gaps and mismatches, and are commonly used to identify orthologs of a newly isolated sequence, according to NCBI.

Methods

The patterns of Stowaway-like transposable element sequences in relation to the S/MARs, pTMX4 and pSR14, were analyzed by examining the results from BLASTn and MEGABLAST searches

for the S/MARs. The searches were preformed based on the flow chart shown below (Figure 1), according to the methods by Delongchamp (2007). All NCBI search parameters were left at default sans the "Max Target Sequence", which was changed to 1000 instead of the default of 100 for both BLASTn and MEGABLAST, and selecting to filter for "low complexity regions" for MEGABLAST (See Figure 1). The number chosen for the max target sequence refers to the highest number of results retrieved by the search according to NCBI. The remaining search parameters are described below were left at default settings. The "expect threshold" is a number that sets a line for determining if the results of the search are statistically significant or by random chance that the sequence would align and result in a hit. The word size, 11 for BLASTn and 28 for MEGABLAST, is the equivalent to a pilot study or search. For the respective search, the word size refers to the exact match of part of the sequence searched to the unknown sequence. Once this match is found, the program continues to try to match more of the sequence. Concerning the scoring parameters, the "match/mismatch" numbers are used to give a total score value associating the first number with a matched nucleotide and the second number with the mismatched number. With the same score value for the BLASTn search, the gap costs also associate a number with either a gap that is penalized to the total score value. Unique to MEGABLAST searches, use a linear gap cost which is also related to match and mismatch values. There are also options to filter some things that may interfere with results. Selecting to filter for "low complexity regions" helps to remove sequences that are not complex which may also mean that the sequences are identified as a match based on coincidence or may not be considered biologically interesting or relevant.

The list of results from all four searches, BLASTn and MEGABLAST for each S/MAR, were copied and saved into a spreadsheet. The results were extracted from the total retrieved results based on specific maximum e-values for the BLASTn searches. An e-value of $n \times 10^{-3}$ was used. Each

location, or record, where either pTMX4 or pSR14 was found in a genome was identified by an accession number, as retrieved through GenBank of the NCBI database portal.

Each record was searched to retrieve the homologous sequence containing the S/MAR, as well as the surrounding sequence information. For each accession record number, the Stowaway-like transposable elements, which had been previously identified, were extracted along with any associated information and placed into a spreadsheet alongside the accession number. The information for each Stowaway-like transposable element was often found buried within the retrieved sequence. These results were analyzed for individual Stowaway-like transposons, the location within the sequence, the length of sequence or fragment, specific Stowaway sequence name if given, order and arrangement, and number of copies.

The Stowaway-like transposable elements were ordered based on the insertion location. Sequence information for each Stowaway-like transposable element was also retrieved: sequence location, sequence length, fragment size (partial vs complete), complement vs non-complement, and the name of the Stowaway or which Stowaway the sequence was most similar. The arrangement which the Stowaway-like transposable elements have inserted was analyzed.

The GenBank result from searching an accession number is displayed in Figure 2. To find the locations of the Stowaway-like transposable elements within the genetic region retrieved from GenBank, first the correct categories had to be located. Box 1a represents the location and names of categories for each genetic element within the genetic region for each accession number. The categories that were further investigated for Stowaways were *mobile elements, repeat regions* and *misc features.* Other DNA elements can be found in categories titled *mRNA, CDS, gene*, etc, but for the purposes of this investigation, these categories were not used. Once a category was identified, detailed information about the category is located to the right of the category title. Box 1b shows the complete identification of a Stowaway-like transposable element. The detailed information first

contains the location within the genetic region, in base pairs. After the location, there is listed information that describes the element. This is the most variable source of information. There is not a standard way to present the information, which makes comparing the genetic elements sometimes difficult. Since there is no standard presentation, it is difficult to describe what information will be presented, how much and in what order. Within this detailed information, it may contain the name of the Stowaway, Polyphemus-3 in Box 1b. It may also contain the type of element, MITE in Box 1b. The information will also describe if the genetic element is a partial (5' or 3') or complete typically. Some other descriptors may also be offered.

The location for each element (Box 1c), as mentioned, is always listed first. The location will not just be the base pair region the element falls in. It will also list "complement" before the base pair region if the element is located on the complement strand (Box 1e). If the word "join" is located before the base pair region, the sequence is not a complete element, it is in pieces that may have insertion mutations over time or have been fragmented for other reasons, like other transposable element insertions (Box 1d).

All the sequence information related to a Stowaway-like transposable element under any of the categories was extracted for each accession number. The information was later organized into several different charts for comparison and conclusions were made about the pattern of insertion for Stowaway-like transposable elements with respect to the S/MARs, pTMX4 and pSR14.

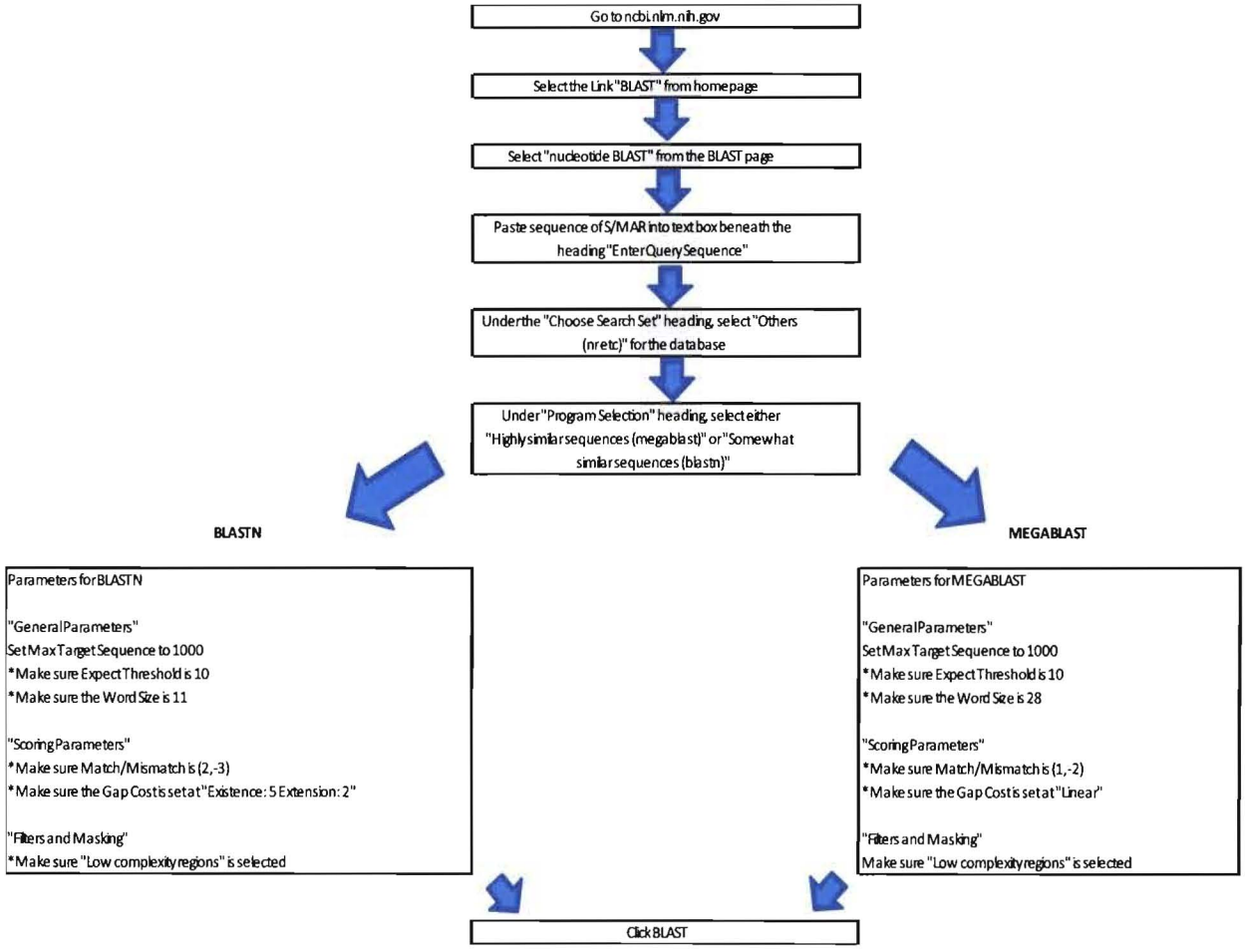


Figure 1. Flow chart instructions for S/MAR BLASTn and MEGABLAST searches

Figure 2. Complete GenBank Results for Accession number EF567062.1, where boxes indicate specific information utilized in collecting data. Box 1a represents the location and names of categories for each genetic element within the genetic region for each accession number. Box 1b shows the complete identification of a Stowaway-like transposable element. Box 1c represents the location for each element. Box 1d shows if the word "join" is located before the base pair region, the sequence is not a complete element, it is in fragmented but identified pieces. Box 1e represents the location of the base pair region the element falls within and indicates if the sequence is on the complement strand or inverted.

Triticum aestivum cultivar Glenlea clone BAC 1648_464 disease resistance protein (Lr1) genomic region

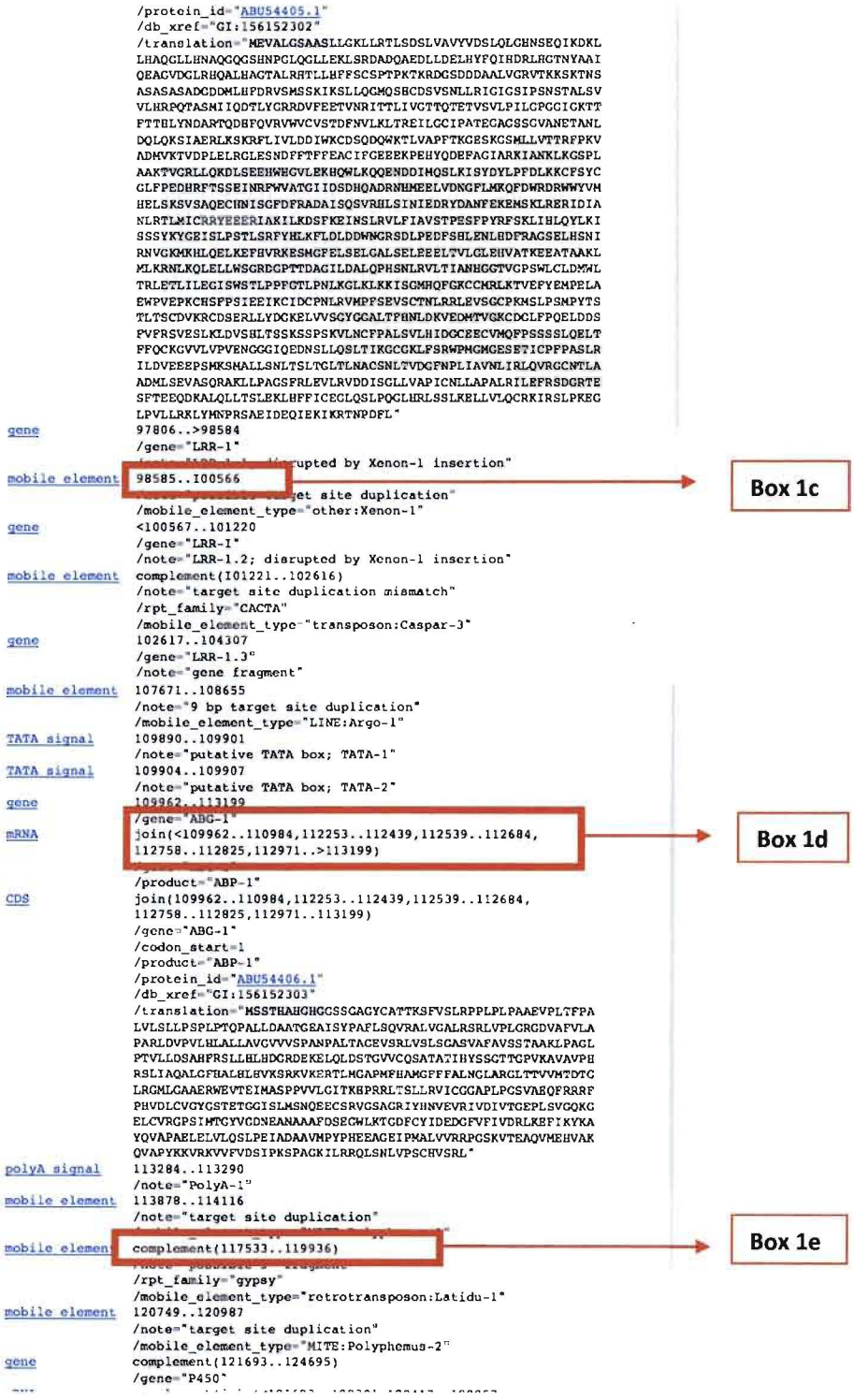
GenBank: EF567062.1

FASTA Graphics

```
LOCUS       EF567062              137614 bp    DNA     linear   PLN 23-AUG-2007
DEFINITION  Triticum aestivum cultivar Glenlea clone BAC 1648_464 disease
            resistance protein (Lr1) genomic region.
ACCESSION   EF567062
VERSION     EF567062.1  GI:156152300
KEYWORDS    .
SOURCE      Triticum aestivum (bread wheat)
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP
            clade; Pooideae; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 137614)
  AUTHORS   Cloutier,S., McCallum,B.D., Loutre,C., Banks,T.W., Wicker,T.,
            Feuillet,C., Keller,B. and Jordan,M.C.
  TITLE     Leaf rust resistance gene Lr1, isolated from bread wheat (Triticum
            aestivum L.) is a member of the large psr567 gene family
  JOURNAL   Plant Mol. Biol. 65 (1-2), 93-106 (2007)
   PUBMED   17611798
REFERENCE   2  (bases 1 to 137614)
  AUTHORS   Cloutier,S., McCallum,B.D., Loutre,C., Banks,T.W., Wicker,T.,
            Feuillet,C., Keller,B. and Jordan,M.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-APR-2007) Agriculture and Agri-Food Canada, Cereal
            Research Centre, 195 Dafoe Road, Winnipeg, MB R3T 2M9, Canada
FEATURES             Location/Qualifiers
     source          1..137614
                     /organism="Triticum aestivum"
                     /mol_type="genomic DNA"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="BAC 1648_464"
     mobile element  1..1846
                     /note="3' fragment"
                     /rpt_family="athila"
                     /mobile_element_type="retrotransposon:Sabrina-3"
     mobile element  1847..5192
                     /note="3' fragment"
                     /rpt_family="gypsy"
                     /mobile_element_type="retrotransposon:Fatima-1"
     mobile element  complement(5193..14010)
                     /note="5' fragment"
                     /rpt_family="athila"
                     /mobile_element_type="retrotransposon:WHAM-1"
     mobile element  complement(14011..15777)
                     /note="5' fragment"
                     /rpt_family="copia"
                     /mobile_element_type="retrotransposon:Angela-2.1"
     mobile element  15778..22531
                     /note="target site duplication"
                     /rpt_family="athila"
                     /mobile_element_type="retrotransposon:Sabrina-1"
     mobile element  complement(22532..24919)
                     /note="5' fragment"
                     /rpt_family="copia"
                     /mobile_element_type="retrotransposon:Angela-2.2"
     mobile element  24920..33544
                     /note="3' fragment"
                     /rpt_family="copia"
                     /mobile_element_type="retrotransposon:Angela-1"
     mobile element  complement(33545..34823)
                     /note="5' fragment"
                     /rpt_family="copia"
                     /mobile_element_type="retrotransposon:Angela-2.3"
     mobile element  34823..34980
                     /note="incomplete element"
                     /mobile_element_type="retrotransposon:Gypsy-1"
     mobile element  38123..39583
                     /mobile_element_type="other:Xabor-1"
     mobile element  50797..52131
                     /mobile_element_type="other:Xusag-1"
     mobile element  53100..53278
                     /note="target site duplication"
                     /mobile_element_type="MITE:Thalos-1"
     mobile element  complement(54267..54362)
                     /note="target site duplication"
                     /mobile_element_type="MITE:Hades-1"
     gene            54812..>58846
                     /gene="Lr1"
     mRNA            54812..>58846
```

Box 1a


mRNA <54812..>58846
/gene="Lr1"
/product="Lr1 protein"
CDS 54812..58846
/gene="Lr1"
/codon_start=1
/product="disease resistance protein"
/protein_id="ABU54404.1"
/db_xref="GI:156152301"
/translation="MAAALGSAATLLGKVFTMLSAAPVAAYVDSLELGHNSQQIRAKL
AHTRGLLHNAQAQVSDVGHNPGLQELLPALSRNADEAEDLLDELHYFQIHDRLHATNY
AATQANFLRHARNALRHTATSSWAACFSCSSAQDDSDSTSGDDELRPHRVIFSRKFKS
VLQDMQTHCDSVSDLLGTIPTSSMPVAVHRPQIGSTIIQDTLYGRRHTFEETVNRIFS
CKHPVSVLPIVGPGGIGKTTFAQHLYNDARTEEHFQVRVWVCVSTDFNVLKLTREILA
CIPATEEGGSSSVANETTNLDHLQRSIVRRLKSKRFLIVLDDIWKCDSQDQWKTLLAP
FTKGETKCSMLLVTTRFPKLAQMMETIDPLELLGLESNDFFTPFEACIFGEDNKPEHF
EDELAGIAQKIADKLKGSPLAAKTVGRLLHKDLSQKHWNGVLEKHQWLKQQNNDDIMP
SLKISYDCLPFDLKKCFSYCGLFPEDHWFTSSEINHFWVAVGIIDSDHQADRNYLEEL
VDNGFLMKKKEYYLDDRCKQKEFDCYVMHDLMHELSKSVSAQECLNISGFDFRADAIP
QSVRHLSINIEDRYDANFEEEMSKLRKKIDIANVRTLMIFREYEEERTAKILKDSFKE
INSLRVLFIVVKSAQSFPDMFSKLIHLQYLKISSPHIDGEMRLPSTLSRFYHLKFLDL
DDWRGSSDLPEDFSHLENLHDFRAESKLHSNIRNVGKMKHLQRLEEFHVKKESMGFEL
SELGPLTELEGGLTVRGLEHVATKEEATAAKLMLKRNLKQLELLWDRDLGGPTTDADI
LDALQPHSNLRVLAIVNHGGTVGPSWLCLDIWLTSLETLTLAGVCWSTLPPFAKLPNL
KGLKLMRISGMHQFGSLCGGTPGKCFMRLKTVEFYEMPELAEWVVESNCHSFPSLEEI
RCRNCPNLRVMPFSEVSFTNLRTLFVSRCPKMSLPSMPHTSTLTDLNVGICDSEGLHY
DGKKLIVIGYGGALASHNLDTVEDMIVERCDGLFPEDLDGSFVFRSVKNLTLHVSRLT
SSKSSSSKVLNCFPALSVLVIVGYEECVMQFPSSSLQKLTFSGCRGLVLVPEEKENG
GGIQEDNSLLQSLTIVGCGKLFSRWPMGMGESETICPFPASLKKLDVFQEPSMKSMAL
LSNLTSLTTLQLNYCSNLTVDGFNPLIAVNLIELQVHRCNTLAADMLSEAASHSQRAK
LLPAGYISRLEKLNVDNNCGLLVAPICNLLAPALHTLVFWIDETMESLTEEQEKALQL
RSAEINEQIEKIKRSNPDLSVSYC"
mobile element 62963..63203
/note="target site duplication"
/mobile_element_type="MITE:Polyphemus-3"
/note="possible target site duplication"
/mobile_element_type="LINE:Karin-1"
mobile element complement(73597..76530)
mobile element 80654..81512
/note="incomplete element"
/rpt_family="CACTA"
/mobile_element_type="transposon:Caspar-5"
mobile element 83761..83894
/note="incomplete element"
/rpt_family="CACTA"
/mobile_element_type="transposon:Caspar-4"
mobile element 85877..88862
/note="possible complete element"
/rpt_family="CACTA"
/mobile_element_type="transposon:Caspar-1"
mobile element 88863..89421
/note="3' fragment"
/rpt_family="CACTA"
/mobile_element_type="transposon:Caspar-2"
gene <89801..>93859
/gene="RGA-1"
mRNA <89801..>93859
/gene="RGA-1"
/product="RGA-1"
CDS 89801..93859
/gene="RGA-1"
/codon_start=1
/product="RGA-1"
/protein_id="ABU54405.1"
Box 1b


/protein_id="ABU54405.1"
/db_xref="GI:156152302"
/translation="MEVALGSAASLLGKLLRTLSDSLVAVYVDSLQLGHNSEQIKDKL
LHAQGLLHNAQGQGSHNPGLQGLLEKLSRDADQAEDLLDELHYFQIHDRLHGTNYAAI
QEACVDGLRHQALHAGTALRHTLLHFFSCSPTPKTKRDGSDDDAALVGRVTKKSKTNS
ASASASADGDDMLHFDRVSMSSKIKSLLQGMQSHCDSVSNLLRIGIGSIPSNSTALSV
VLHRPQTASMIIQDTLYGRRDVFEETVNRITTLIVGTTQTETVSVLPILGPGGIGKTT
FTTBLYNDARTQDHFQVRVWVCVSTDFNVLKLTREILGCIPATEGAGSSGVANETANL
DQLQKSIAERLKSKRFLIVLDDIWKCDSQDQWKTLVAPFTKGESKGSMLLVTTRFPKV
ADMVKTVDPLELRGLESNDFFTFFEACIFGEEEKPEHYQDEFAGIARKIANKLKGSPL
AAKTVGRLLQKDLSEEHWEGVLEKHQWLKQQENDDIMQSLKISYDYLPFDLKKCFSYC
GLFPEDHRFTSSEINRFWVATGIIDSDHQADRNHMEELVDNGFLMKQFDWRDRWWYVH
HELSKSVSAQECHNISGFDFRADAISQSVRHLSINIEDRYDANFEKEMSKLRERIDIA
NLRTLMICRAYEEERIAKILKDSFKEINSLRVLFIAVSTPESFPYRFSKLIHLQYLKI
SSSYKYGEISLPSTLSRFYHLKFLDLDDWNGRSDLPEDFSHLENLEDFRAGSELHSNI
RNVGKMKHLQELKEFHVRKESMGFELSELGALSELEEELTVLGLEHVATKEEATAAKL
MLKRNLKQLELLWSGRDGPTTDAGILDALQPHSNLRVLTIANHGGTVGPSWLCLDMWL
TRLETLILEGISWSTLPPFGTLPNLKGLKLKKISGMHQFGKCCMRLKTVEFYEMPELA
EWPVEPKCHSFPSIEEIKCIDCPNLRVMPFSEVSCTNLRRLEVSGCPKMSLPSMPYTS
TLTSCDVKRCDSERLLYDGKELVVSGYGGALTPHNLDKVEDMTVGKCDGLFPQELDDS
PVFRSVESLKLDVSHLTSSKSSPSKVLNCFPALSVLHIDGCEECVMQFPSSSSLQELT
FFQCKGVVLVPVENGGGIQEDNSLLQSLTIKGCGKLFSRWPMGMGESETICPFPASLR
ILDVEEEPSMKSMALLSNLTSLTGLTLNACSNLTVDGFNPLIAVNLIRLQVRGCNTLA
ADMLSEVASQRAKLLPAGSFRLEVLRVDDISGLLVAPICNLLAPALRILEFRSDGRTE
SFTEEQDKALQLLTSLEKLHFFICEGLQSLPQGLHRLSSLKELLVLQCRKIRSLPKEG
LPVLLRKLYMNPRSAEIDEQIEKIKRTNPDFL"
gene
97806..>98584
/gene="LRR-1"
mobile element
98585..100566
/mobile_element_type="other:Xenon-1"
gene
<100567..101220
/gene="LRR-1"
/note="LRR-1.2; disrupted by Xenon-1 insertion"
mobile element
complement(101221..102616)
/note="target site duplication mismatch"
/rpt_family="CACTA"
/mobile_element_type="transposon:Caspar-3"
gene
102617..104307
/gene="LRR-1.3"
/note="gene fragment"
mobile element
107671..108655
/note="9 bp target site duplication"
/mobile_element_type="LINE:Argo-1"
TATA signal
109890..109901
/note="putative TATA box; TATA-1"
TATA signal
109904..109907
/note="putative TATA box; TATA-2"
gene
109962..113199
/gene="ABG-1"
mRNA
join(<109962..110984,112253..112439,112539..112684,
112758..112825,112971..>113199)
/product="ABP-1"
CDS
join(109962..110984,112253..112439,112539..112684,
112758..112825,112971..113199)
/gene="ABG-1"
/codon_start=1
/product="ABP-1"
/protein_id="ABU54406.1"
/db_xref="GI:156152303"
/translation="MSSTHAHGHGCSSCAGYCATTKSFVSLRPPLPLPAAEVPLTFPA
LVLSLLPSPLPTQPALLDAATGEAISYPAFLSQVRALVGALRSRLVPLGRGDVAFVLA
PARLDVPVLHLALLAVGVVVSPANPALTAGEVSRLVSLSGASVAFAVSSTAAKLPAGL
PTVLLDSAHFRSLLHLHDCRDEKELQLDSTGVVCQSATATIHYSSGTTGPVKAVAVPH
RSLIAQALGFHALHLHVKSRKVKERTLHGAPMFHAMGFFFALNGLARGLTTVVMTDTG
LRGMLGAAERWEVTEIMASPPVVLGITKHPRRLTSLLRVICGGAPLPGSVAEQFRRRF
PHVDLCVGYGSTETGGISLMSNQEECSRVGSAGRIYHNVEVRIVDIVTGEPLSVGQKG
ELCVRGPSIMTGYVGDNEANAAAFDSEGWLKTGDFCYIDEDGFVFIVDRLKEFIKYKA
YQVAPAELELVLQSLPEIADAAVHPYPHEEAGEIPMALVVRRPGSKVTEAQVMEHVAK
QVAPYKKVRKVVFVDSIPKSPAGKILRRQLSNLVPSCHVSRL"
polyA signal
113284..113290
/note="PolyA-1"
mobile element
113878..114116
/note="target site duplication"
mobile element
complement(117533..119936)
/rpt_family="gypsy"
/mobile_element_type="retrotransposon:Latidu-1"
mobile element
120749..120987
/note="target site duplication"
/mobile_element_type="MITE:Polyphemus-2"
gene
complement(121693..124695)
/gene="P450"
Box 1c
Box 1d
Box 1e

```
                /gene="P450"
mRNA            complement(join(<121693..122301,122417..122857,
                124201..>124695))
                /gene="P450"
                /product="P450"
CDS             complement(join(121693..122301,122417..122857,
                124201..124695))
                /gene="P450"
                /codon_start=1
                /product="P450"
                /protein_id="ABU54407.1"
                /db_xref="GI:156152304"
                /translation="MLITHTTLLLSLIQQWELIFLALASVVSLLLLTRSRSRSSQKEM
                KLPPGPAPVPFLGNLHQLGRLPYRTLRDLARLHGPVMQLHLGKAPTVVLSSADAAWEG
                LKVHDLDCCTRPVSPGPKRLTYDLKNVAFAPFGSYWREIRKLLVVELLSGRRVKAAWY
                ARQEQVEKLISTLSRAGGKPMALDEHILSLSDGIIGTVAFGNIYGSDKFTQNNNFQHA
                LDEAAEVLSSSSAEDFFPKAIGWLIDRITGAFARRERIFKQLDAFFEMVIEQHLDPKR
                TKPENGGDLVDVLIAHWKEHRGTLKFTKDHVKALIFDTFIAGVDTSSVTILWAMSELM
                RSPRVLSKVQAEIRALVGGNDRVRSEDVSKLEYLKLVVKETLRLHPPAPLLVPRETMR
                HIQIGGYDVPAKTRIYVNAWAIGRDPASWPDNPEEFNPERFEANEIDFKGEHPQMLPF
                GTGRRICPGISMGMATVEFTLANMLCCFQWALPDGMVPQDVCMEEEGKINFHRKTPLV
                LVPTIR"
mobile_element  complement(127912..128400)
                /note="3' fragment"
                /rpt_family="athila-like"
                /mobile_element_type="retrotransposon:Sabrina-2.2"
mobile_element  128401..129550
                /note="target site duplication"
                /rpt_family="CACTA"
                /mobile_element_type="transposon:TAT-1"
mobile_element  complement(129551..131257)
                /note="3' fragment"
                /rpt_family="athila-like"
                /mobile_element_type="retrotransposon:Sabrina-2.2"
mobile_element  131302..135543
                /note="incomplete element"
                /rpt_family="CACTA"
                /mobile_element_type="transposon:TAT-2"
mobile_element  complement(135544..136981)
                /note="3' fragment"
                /rpt_family="gypsy"
                /mobile_element_type="retrotransposon:Daniela-1"
```

CHAPTER 3

RESULTS

Analysis of pTMX4

Overall Results

BLASTn

Of the 219 sequences identified to be homologous to the sequence of pTMX4 by the BLASTn search parameters, 136 sequences were extracted due to the parameters of an e-value equal to or less than 1.00×10^{-3}. Of the 136 extracted sequences, 57 sequences were identified to contain a Stowaways or potential Stowaways. Forty-six of the 57 sequences contained identified and fully annotated Stowaways. These 46 sequences were compared to analyze the relationship between S/MARs and Stowaway transposable elements (Appendix A-1).

MEGABlast

A total of the 106 were sequences identified to be homologous to pTMX4 by the MEGABlast search parameters. These sequences were extracted utilizing a maximum e-value equal to or less than 1.00×10^{-3}. Of the 106 extracted sequences, 55 sequences were identified to contain a Stowaway or potential Stowaway sequence. Forty-three of these 55 sequences contained identified and fully annotated Stowaways. These 43 sequences were compared to analyze the relationship between S/MARs and Stowaway transposable elements (Appendix A-1).

Species identified

BLASTn

Seventeen different species were identified by the BLASTn search to have a sequence homologous to pTMX4 (Appendix A-2). These include species from the 10 genera: *Aegilops, Avena,*

Bromus, Dasypyrum, Hordeum, Leymus, Puccinia, Secale, Thinopyrum and *Triticum.* Of the 136 sequences identified by the BLASTn search to be homologous to pTMX4, 69 sequences were from the genus *Triticum*, the most results of all genera; 12 were from *Aegilops;* 10 were from *Puccinia*; and 31 were from *Hordeum*. The remaining genera each contained fewer than 10 sequences homologous to pTMX4. Only 5 of the 17 species contained sequences that were suggestive of the presence of Stowaways. These species came from the genera *Aegilops, Hordeum,* and *Triticum.* These same 5 species contained slightly fewer sequences containing fully annotated Stowaways.

Of the 69 sequences from *Triticum*, 44 sequences contained sequence information suggesting the presence of Stowaways. Of these 44 sequences, 37 contained fully annotated Stowaways. Of the 31 sequences from *Hordeum,* 10 sequences contained information suggesting the presence of Stowaways. Seven of the 10 sequences contained fully annotated Stowaways. In *Aegilops,* 3 of the 12 sequences identified contained information for Stowaways and/or potential Stowaways, where only 2 of the 3 sequences contained fully annotated Stowaways.

MEGABlast

Nine different species were identified by the MEGABlast search to have a sequence homologous to pTMX4 (Appendix A-2). These include species from the 6 genera: *Aegilops, Hordeum, Leymus, Puccinia, Secale,* and *Triticum.* Of the 106 sequences identified by the MEGABlast search to be homologous to pTMX4, 65 sequences were from the genus *Triticum,* the most results of all genera; 10 were from *Aegilops;* 9 were from *Puccinia*; and 19 were from *Hordeum*. The remaining genera each contained fewer than 10 sequences homologous to pTMX4. Only 5 of the 9 species contained sequences that were suggestive of the presence of Stowaways. These species came from the genera *Aegilops, Hordeum,* and *Triticum.* These same 5 species contained slightly fewer sequences containing fully annotated Stowaways. These are the same species identified by the BLASTn search.

Of the 65 sequences from *Triticum*, 43 sequences contained sequence information suggesting the presence of Stowaways. Of these 43 sequences, 36 contained fully annotated Stowaways. Of the 19 sequences from *Hordeum*, 9 sequences contained information suggesting the presence of Stowaways. Five of the 9 sequences contained fully annotated Stowaways. In *Aegilops,* 3 of the 10 sequences identified contained information for Stowaways and/or potential Stowaways, where only 2 of the 3 contained fully annotated Stowaways.

Stowaway Relationships

S/MAR from BLASTn and MEGABlast

The results of the BLASTn and MEGABlast searches were individually organized by e-value, as received from the search. The e-values were associated with the levels of homology to the S/MAR pTMX4, as the e-values increased the levels of homology decreased. A Stowaway pattern became visible as the e-values increased (levels of homology decreased) with respect to the sequences homologous to pTMX4 (Appendix B-3, B-9). As the level of homology decreased, the quantity and variety of Stowaway sequences identified decrease as well.

Species from BLASTn and MEGABlast

Five species were identified to contain Stowaways from the BLASTn and MEGABlast searches: *Aegilops tauschii, Hordeum vulgare, Triticum aestivum, Triticum monococcum,* and *Triticum turgidum* (Appendix B-4, B-10). The 2 sequences identified from *Aegilops tauschii* from both BLASTn and MEGABlast searches contained 4 different Stowaways. The 7 sequences identified from the BLASTn search, 5 sequences from the MEGABlast search, of *Hordeum vulgare* contained 6 different Stowaways. Within the *Triticum genus*, *monococcum* contains 9 different Stowaways across 6 sequences identified in both searches. *T. monococcum* had the same approximate quantity of individual Stowaways per sequence as *Hordeum vulgare*. These species had no more than 7 Stowaways or Stowaway fragments per sequence. Just shy of doubling this quantity, *Triticum*

turgidum and *Aegilops tauschii* have no more than 12 Stowaways or Stowaway fragments per sequence. *Triticum turgidum* contained 13 different Stowaways, many Stowaways occurring once, found in 8 sequences identified in both searches. Twenty-two different Stowaways were found within sequences identified from *Triticum aestivum*. These were identified within 23 sequences from the BLASTn search and 22 sequences from the MEGABlast search. These sequences had a quantity of Stowaways triple the other species, with no more than 67 Stowaways per sequence. There does not appear to be a pattern within each species based on the levels of homology to pTMX4 in this analysis.

Quantity from BLASTn and MEGABlast

Out of the 46 BLASTn search sequences with fully annotated Stowaways, the distribution of full Stowaway sequences versus BLASTn sequences followed an exponential relationship (Appendix A-5, B-5). A single Stowaway was identified in 3 sequences. There were 2 Stowaways in five sequences, and 3 in five other sequences. Four Stowaways were found in each of 4 sequences. The most common quantity of Stowaways was 5, where 8 sequences were found to have 5. Only two sequences had 6 Stowaways, and three had 7. A total of 8 Stowaways were only found in single sequence. Six sequences have 10 Stowaways each and 1 sequence has 11 Stowaways. The increase in Stowaway quantity was slow, increasing by only a few Stowaways per sequence until the quantity of Stowaways within one sequence exceeded 12. This number quickly jumped to 17 Sequences for each of two sequences. After this, there were 28 Stowaways within a sequence. Five sequences which had the most fully identified Stowaways contained 32, 40, 45, and 67. The MEGABlast search revealed similar results, with the exception of 1 sequence with 1 Stowaway, 1 sequence with 4 Stowaways, and 1 sequence with 7 Stowaways.

Stowaways Identified

The 12 Stowaways identified within the 46 BLASTn sequences and 43 MEGABlast sequence homologous to pTMX4 were as follows: Thalos, Icarus, Athos, Damocles, Hades, Polyphemus, Pan, Eos, Jason, Stolos, Fortuna, and Oleus. These were identified due to at least one occurrence of an annotation including the descriptor "Stowaway" alongside the name (Appendix A-4). Twelve transposable elements that might be Stowaway-like associated with sequences homologous to pTMX4 were also identified, they included: Tantalos, Nisos, Orpheus, Antonio, Phoebus, Yabi, Revo, Ganesh, Boulos, Belus, Charon, and Gorgon. These potential Stowaway-like transposable elements were found less than five times across all 46 sequences (Appendix A-4). These sequences did not contain an annotation descriptor "Stowaway" alongside the name of the element.

Stowaway Relationship to other Stowaways

Several unique trends were revealed among the individual Stowaways (Appendix B-6, B-7, B-8, B-12, B-13, B-14). Athos occurred in multiples on several occasions in sequences with the accession numbers: FN564434.1, FN564430.1, FN564427.1, FN564428.1, FN645450.1, CT009625.1, EF067844.1, EU835198.1, and DQ871219.1. In these multiples, Athos was often truncated or fragmented.

Several occasions show Thalos as a triplet. These sequences had accession numbers: FN564431.1, FN645450.1, and DQ249273.1. The majority of the time, Thalos was found as a singlet. One unique occurrence in accession number AY146588.1 showed this pattern twice: Thalos fragment, a complete Thalos, and another Thalos fragment. Thalos most often occurred downstream from an Athos.

Hades appeared to occur most often by itself. In 3 of the 7 occurrences of Oleus, Oleus was paired with itself. In 2 of these 3 occurrences, the downstream Oleus is inverted. Fortuna also tended to be a singlet. On the other hand, Polyphemus tended to occur in multiples or within a

distance of 1 or 2 other Stowaways separating the parts of the multiple. Pan occurred often near Polyphemus, more often downstream, and is usually a singlet. Of the three sequences with only one identified Stowaway, 2 of the sequences had only Icarus and 1 contained only Damocles.

Several occurrences of an adjacent pair of Thalos and Icarus were identified where one or both were fragmented. This occurred in sequences with accession numbers: DQ900687.1, AY146587.2, CR626926.1, CT009587.1, and FN564434.1. Even more sequences contained this pattern with an additional Stowaway between the Icarus and Thalos or contained non-fragmented Thalos and Icarus adjacent to each other. When paired with another Stowaway, the additional sequence was often Athos, which was found both as a complete or fragmented sequence.

Other unique patterns identified occurred when the Stowaways fell in the same or similar pattern between sequences. Sequences with accession numbers FN564428.1 and FN645450.1 had a pattern in which the Stowaways were found in the following order: Thalos, Hades, Athos, Icarus, Icarus, Thalos, Thalos. While other Stowaways were identified within this arrangement, the general pattern remained consistent.

A pattern of Hades, Icarus, Athos was found in other sequences: accession numbers CT009587.1, FN564429.1 (with an inverted Oleus as well); FN564426.1 (with an inverted Polyphemus); and FN564434.1 (reversed order). Similarly, the pattern Hades, Athos, Icarus was also present. The sequences with this pattern had accession numbers FN564429.1, FN564428.1, FN564428.1 (with a Thalos), FN564434.1 (also with a Thalos, reversed order).

Between the sequences with accession numbers CT009588.1 and CT009585.1, a pattern of Stowaways was present as follows: inverted Icarus, inverted Thalos, inverted and fragmented Jason, inverted and fragmented Hades, inverted Damocles, and Hades. In CT009588.1, an inverted, fragmented Fortuna was present at the start of the arrangement and the final Hades was a 5' fragment.

Analysis of pSR14

Overall Results

BLASTn

A total of 1000 sequences were identified to be homologous to the sequence of pSR14 by the BLASTn search parameters, where the total list was truncated after 1000 matches, as determined by parameters discussed in Chapter 2 (Figure 1). Of these 1000 sequences, 973 sequences were extracted due to the parameters of an e-value equal to or less than 1.00×10^{-3}. Of the 973 extracted sequences, 75 sequences were identified to contain a Stowaways or potential Stowaways. Fifty of the 75 sequences contained identified and fully annotated Stowaways. These 50 sequences were compared to analyze the relationship between S/MARs and Stowaway transposable elements (Appendix A-1).

MEGABlast

A total of the 45 were sequences identified to be homologous to pSR14 by the MEGABlast search parameters. These sequences were extracted utilizing a maximum e-value equal to or less than 1.00×10^{-3}. Of the 45 extracted sequences, 5 sequences were identified to contain a Stowaway or potential Stowaway sequence. Three of these 5 sequences contained identified and fully annotated Stowaways. These 3 sequences were compared to analyze the relationship between S/MARs and Stowaway transposable elements (Appendix A-1).

Species identified

BLASTn

Forty-two different species were identified by the BLASTn search to have a sequence homologous to pSR14 (Appendix A-3). These include species from the 20 genera: *Aegilops, Brachypodium, Equus, Festuca, Homo, Hordeum, Lolium, Medicago, Mus, Oryza, Pan, Panicum, Plasmodium, Puccinia, Saccharum, Secale, Sorghum, Sus, Triticum,* and *Zebrafish.* Of the 973

sequences identified by the BLASTn search to be homologous to pSR14, 729 sequences were from the genus *Oryza*, the most results of all genera; 98 were from *Triticum;* 12 were from *Zebrafish;* 18 were from *Sorghum;* 39 were from *Hordeum;* 23 were from *Mus;* and 18 were from *Aegilops*. The remaining genera each contained fewer than 10 sequences homologous to pSR14. Only 11 of the 42 species contained sequences that were suggestive of the presence of Stowaways. These species came from the genera *Aegilops, Homo, Hordeum, Oryza, Sorghum,* and *Triticum.* Of the11 species, only 5 species had sequences which contained fully annotated Stowaways. These species were from the genera *Aegilops, Hordeum,* and *Triticum*.

Of the 98 sequences from *Triticum*, 40 sequences contained sequence information suggesting the presence of Stowaways. Of these 40 sequences, 34 contained fully annotated Stowaways. Of the 39 sequences from *Hordeum*, 16 sequences contained information suggesting the presence of Stowaways. Fourteen of the 16 sequences contained fully annotated Stowaways. In *Aegilops,* 2 of the 18 sequences identified contained information for Stowaways and/or potential Stowaways. The same 2 sequences contained fully annotated Stowaways.

MEGABlast

Seven different species were identified by the MEGABlast search to have a sequence homologous to pSR14 (Appendix A-3). These include species from the 4 genera: *Brachypodium, Festuca, Oryza,* and *Triticum.* Of the 45 sequences identified by the MEGABlast search to be homologous to pSR14, 32 sequences were from the genus *Oryza,* the most results of all genera. The remaining genera each contained fewer than 10 sequences homologous to pSR14. Only 2 of the 7 species contained sequences that were suggestive of the presence of Stowaways. These species were both from the genus *Triticum.* Only one species, *Triticum aestivum,* contained fully annotated Stowaways in the sequences identified by this search.

Stowaway Relationships

S/MAR from BLASTn and MEGABlast

The results of the BLASTn and MEGABlast searches were individually organized by e-value, as received from the search. The e-values were associated with the levels of homology to the S/MAR pSR14, as the e-values increased the levels of homology decreased. A Stowaway pattern became visible as the e-values increased (levels of homology decreased) with respect to the sequences homologous to pSR14 (Appendix C-3, C-9). As the level of homology decreased, the quantity and variety of Stowaway sequences identified decrease as well.

Species from BLASTn and MEGABlast

Five species were identified to contain Stowaways from the BLASTn search: *Aegilops tauschii, Hordeum vulgare, Triticum aestivum, Triticum monococcum,* and *Triticum turgidum* (Appendix C-4, C-10). The 2 sequences identified from *Aegilops tauschii* from the BLASTn search contained 4 different Stowaways. The 14 sequences identified from the BLASTn search of *Hordeum vulgare* contained 11 different Stowaways. Within the *Triticum genus*, *monococcum* contains 6 different Stowaways across 4 sequences identified in both searches. *Triticum turgidum* contained 13 different Stowaways, many Stowaways occurring once, found in 9 sequences identified in the BLASTn search. Twenty-one different Stowaways were found within sequences identified from *Triticum aestivum.* These were identified within 21 sequences. *T. monococcum, T. turgidum, A. tauschii,* and *H. vulgare* all had the same approximate quantity of individual Stowaways per sequence. These species had Stowaways which ranged from 1 per sequence up to 13 Stowaways or Stowaway fragments per sequence. Tripling this quantity, *Triticum aestivum* had no more than 69 Stowaways per sequence, containing the highest quantity and most variation of Stowaways between the species. The MEGABlast search revealed 3 sequences all from the same species,

Triticum aestivum. The sequences contained 14 different Stowaways with a range of 4 to 69 Stowaways or Stowaway fragments per sequence.

Quantity from BLASTn and MEGABlast

Out of the 50 BLASTn search sequences with fully annotated Stowaways, the distribution of fully annotated Stowaways or Stowaway fragments versus BLASTn sequences followed an exponential relationship (Appendix A-6, C-5). A single Stowaway was identified in 6 sequences. There were 2 Stowaways in four sequences, and 3 in six other sequences. Four Stowaways were found in each of 3 sequences. Five Stowaways were identified in 5 sequences. Four sequences had 6 Stowaways, and three had 7. Five different sequences had 10 Stowaways each and 2 sequences had 11 Stowaways. Only 1 sequence contained 12 Stowaways and 3 sequences were found to have 13 Stowaways. The increase in Stowaway quantity was slight quicker than pTMX4, increasing still by a few Stowaways per sequence until the quantity of Stowaways within one sequence exceeded 13. This number quickly jumped to 17 Stowaways for each of two sequences. After this, there were 22 Stowaways within a sequence. Five sequences which had the most fully identified Stowaways contained 28, 34, 38, 45, and 69. The MEGABlast search revealed quite different results, with one sequence containing 4, one containing 45 and one containing 69 Stowaways or Stowaway fragments.

Stowaways Identified

The 12 Stowaways identified within the 50 BLASTn sequences and 3 MEGABlast sequences homologous to pSR14 were as follows: Athos, Thalos, Icarus, Hades, Polyphemus, Pan, Damocles, Oleus, Stolos, Fortuna, Eos, and Jason. These were identified due to at least one occurrence of an annotation including the descriptor "Stowaway" alongside the name (Appendix A-4). Twelve transposable elements that might be Stowaway-like associated with sequences homologous to pSR14 were also identified, they included: Aison, Nisos, Antonio, Yabi, Revo, Ganesh, Phoebus,

Tantalos, Nyx, Midas, Xenos, and Gorgon. These potential Stowaway-like transposable elements were found less than five times across all 50 sequences (Appendix A-4). These sequences did not contain an annotation descriptor "Stowaway" alongside the name of the element.

Stowaway Relationship to other Stowaways

Several unique trends were revealed among the individual Stowaways (Appendix C-6, C-7, C-8, C-12, C-13, C-14). Athos occurred in multiples on several occasions in sequences with the accession numbers: FN564434.1, FN564428.1, FN645450.1, CT009625.1, CT009735.1, AY853252.1, EU835198.1, EF067844.1, and AY951945.1. In these multiples, Athos was often truncated or fragmented.

Several occasions showed Thalos as a triplet. These sequences had accession numbers: FN564431.1, DQ871219.1, AY661558.1, and DQ249273.1. Thalos was also found in a triplet that was slightly separated: FN564428.1 (with an Athos, second triplet with an Icarus) and FN564434.1 (with 2 Athos). The majority of the time, Thalos was found as a singlet as with pTMX4. One unique occurrence in accession number AY146588.1 showed this pattern twice: Thalos fragment, a complete Thalos, and another Thalos fragment. Thalos most often occurred downstream from an Athos.

Hades appeared to occur approximately half the time as a doublet, the other half was as a singlet. Hades was also often found fragmented. In 2 of the 8 occurrences of Oleus, Oleus was paired with itself. Fortuna also tended to be a singlet. Polyphemus occasionally occurred in doublets, but was more often found as a singlet in pSR14. Many of the singlets of Polyphemus were separated by a distance of 1 or 2 Stowaways similar to pTMX4. Pan occurred often near Polyphemus, more often downstream, and usually as a singlet. Of the four sequences with only one identified Stowaway, 2 of the sequences had only Icarus, 1 had only Thalos, and 1 contained only Damocles.

Several occurrences of an adjacent pair of Thalos and Icarus (either order) were identified; occasionally one or both were fragmented. The pair in which one or both were fragmented occurred in sequences with accession numbers: DQ900687.1, CR626926.1, CT009625.1, and CT009587.1. Occurrences of a non-fragmented Thalos-Icarus pair were found within most sequences identified. Additional occurrences of this pattern included a Stowaway between the Icarus and Thalos. When paired with another Stowaway, the additional sequence was often Athos, either as a complete or fragmented sequence.

Unusual patterns identified occurred when the Stowaways fell in the same or similar pattern among a few sequences. Sequences with accession numbers FN564428.1 and FN645450.1 had a pattern in which the Stowaways were found in the following order: Thalos, Hades, Athos, Icarus, Icarus, Thalos, Thalos. While other Stowaways were identified within this arrangement, the general pattern remained consistent.

A pattern of Hades, Icarus, Athos was found in other sequences: accession numbers CT009587.1, FN564429.1 (2 occurrences, one with an inverted Oleus); FN564426.1 (with an inverted Polyphemus); FN645450.1; CT009586.1; CR626929.1; and CR626933.1 (only 3 Stowaways present). Similarly, the pattern Hades, Athos, Icarus was also present. The sequences with this pattern had accession numbers FN564428.1, FN564428.1 (with a Thalos), CT009586.1 (reversed order), and CR626929.1 (reversed order).

Between the sequences with accession numbers CT009588.1 and CT009585.1, a pattern of Stowaways was present as follows: inverted Icarus, inverted Thalos, inverted and fragmented Jason, inverted and fragmented Hades, inverted Damocles, and Hades. In CT009588.1, an inverted, fragmented Fortuna was present at the start of the arrangement and the final Hades was a 5' fragment. Majority of the patterns were analyzed with the BLASTn results of pSR14 due to the higher quantity of results as compared to the MEGABlast results of pSR14.

Comparison of pTMX4 and pSR14 Results

BLASTn Results

The sequence hits achieved for pTMX4 did not reach capacity, as determined by the max target sequence, as compared to pSR14. This resulted in 771 additional sequences identified using pSR14 versus pTMX4. Sequences which had e-values less than or equal to the 1.00×10^{-3} parameter were also higher in the results from pSR14. The total difference between the sequence counts was 837 sequences. The difference was greater between these sequences (837), which had no more than the maximum e-value, than the difference between the total sequence counts (771). The difference between the pTMX4 and pSR14 sequences containing Stowaways or potential Stowaways was significantly less, with a difference of only 18 sequences. This number continued to decrease once sequences with only fully annotated Stowaways were counted, resulting in a difference of only 4 sequences. In all occurrences, pSR14 had the greater quantity of sequences (Appendix A-1).

MEGABlast Results

The MEGABlast results for pTMX4 and pSR14 were significantly smaller than the BLASTn results. Total sequences were 106 using pTMX4 and 45 using pSR14. This showed a difference of 61 additional sequences identified using pTMX4 versus pSR14. Sequences which had e-values less than or equal to the 1.00×10^{-3} parameter were the same quantities for both S/MARs, thus the difference was also the same. The difference between the pTMX4 and pSR14 sequences containing Stowaways or potential Stowaways was slightly less, with a difference of 52 sequences. This number continued to decrease once sequences with only fully annotated Stowaways were counted, resulting in a difference of 40 sequences. In all occurrences, pTMX4 had the greater quantity of sequences (Appendix A-1).

Species Identified

BLASTn

Of the 50 total species, 42 species were identified to contain pSR14 and 17 were identified to contain pTMX4 using the BLASTn results (Appendix A-2, A-3). There were 9 species in common between the S/MARs from 5 genera which included: *Aegilops, Hordeum, Puccinia, Secale,* and *Triticum.* The 15 genera unique to pSR14 included: *Brachypodium, Equus, Festuca, Homo, Lolium, Medicago, Mus, Orzya, Pan, Panicum, Plasmodium, Saccharum, Sorghum, Sus,* and *Zebrafish.* The 5 genera unique to pTMX4 included: *Avena, Bromus, Dasypyrum, Leymus,* and *Thinopyrum.*

Eighteen sequences were identified in *Aegilops* using pSR14 and 12 sequences using pTMX4 with a difference of 6 sequences. Thirty-nine sequences were identified in *Hordeum* using pSR14 and 31 sequences using pTMX4 with a difference of 8 sequences. Only 1 sequence was identified in *Puccinia* using pSR14 whereas 10 were identified using pTMX4, resulting in a difference of 9. Seven sequences were identified for the genus *Secale* from each S/MAR. Ninety-eight sequences were identified in *Triticum* using pSR14 whereas only 69 were identified with pTMX4 resulting in a difference of 29 sequences.

Sequences which contained Stowaways or potential Stowaways were found in only 5 species, 3 genera. All 5 species were the same between the S/MARs which included: *Aegilops tauschii, Hordeum vulgare, Triticum aestivum, Triticum monococcum,* and *Triticum turgidum*. Only 2 sequences were found in *A. tauschii* using pSR14 and only 3 sequences using pTMX4, with a difference of 1. Sixteen sequences were identified in *H. vulgare* using pSR14 and 10 using pTMX4 with a difference of 6. *T. aestivum* contained 24 sequences identified using pSR14 and 28 sequences identified using pTMX4, difference of 4. *T. monococcum* contained 6 sequences identified using pSR14 and 7 sequences identified using pTMX4, difference of 1. *T. turgidum* contained 10 sequences identified using pSR14 and 9 sequences identified using pTMX4, also a difference of 1.

Sequences which contained fully annotated Stowaways were found in the same 5 species using pTMX4 and pSR14. *A. tauschii* had 2 sequences with fully annotated Stowaways for each S/MAR. There were twice as many sequences, 14 using pSR14 than there were using pTMX4, 7 in *H. vulgare.* Twenty-three sequences were identified using pTMX4 and 21 using pSR14 in *T. aestivum*, with a difference of 2. Six sequences were identified using pTMX4 and 4 using pSR14 in *T. monococcum*, with a difference of 2 as well. *T. turgidum* had 8 sequences identified with pTMX4 and 9 identified with pSR14, with a difference of only 1.

MEGABlast

Of the 13 total species, 7 species were identified to contain pSR14 and 9 were identified to contain pTMX4 using the BLASTn results (Appendix A-2, A-3). There were 3 species in common between the S/MARs all from the genus *Triticum.* The 3 genera unique to pSR14 included: *Brachypodium, Festuca, Orzya.* The 5 genera unique to pTMX4 included: *Aegilops, Hordeum, Leymus, Puccinia,* and *Secale.*

The 3 species in common were *T. aestivum, T. monococcum,* and *T. turgidum*. Sequences identified in *T. aestivum* were 40 using pTMX4 and 8 using pSR14, resulting in a difference of 32. Sequences identified in *T. monococcum* were 8 using pTMX4 and 1 using pSR14, resulting in a difference of 7. Sequences identified in *T. turgidum* were 14 using pTMX4 and 1 using pSR14, resulting in a difference of 13.

Sequences which contained Stowaways or potential Stowaways were found in only 2 common species, *Triticum aestivum* and *Triticum monococcum*. Sequences identified in *T. aestivum* were 27 using pTMX4 and 3 using pSR14, resulting in a difference of 24. Sequences identified in *T. monococcum* were 7 using pTMX4 and 1 using pSR14, resulting in a difference of 6.

There was only one species in common which contained sequences with fully annotated Stowaways, *Triticum aestivum.* There were 22 sequences identified using pTMX4, but only 3 sequences identified using pSR14, resulting in a difference of 19 sequences.

Stowaway Relationships

S/MARs

The same general relationship between the S/MAR homology and the quantity and variety of Stowaways was observed in both pTMX4 and pSR14 (Appendix B-3, B-9, C-3, C-9).

Species

Both S/MARs had the highest quantity and variety of Stowaways when in the species *T. aestivum. T. monococcum* and *H. vulgare* had an increased amount of Stowaways within sequences identified using pSR14. *T. turgidum* and *A. tauschii* had similar quantities and varieties in sequences identified with both S/MARs (Appendix B-4, B-10, C-4, C-10).

Quantity

The highest number of Stowaways per sequence identified by pSR14 was 69 and 67 for pTMX4, a difference of only 2. The lowest for each was 1 Stowaway. The same general exponential relationship of the Stowaway quantity was represented in each S/MAR. The only difference appeared to be the number of Stowaways per sequence where the more drastic increase occurred was 12 with pTMX4 and 13 with pSR14 (Appendix A-6, B-5, C-5).

Stowaways Identified

All 12 of the Stowaways were consistent between the two S/MARs. The only differences existed in the potential Stowaways (Appendix A-4). The potential Stowaways that were common between the pTMX4 and pSR14 included: Tantalos, Antonio, Yabi, Revo, Ganesh, Phoebus, Gorgon, and Nisos. Potential Stowaways that were unique to pTMX4 included: Orpheus, Boulos, Belus, and Charon. Potential Stowaways that were unique to pSR14 included: Aison, Nyx, Midas, and Xenos.

Stowaway Relationship to other Stowaways

A majority of the Stowaway patterns observed in the sequences identified using pTMX4 and pSR14 were the same or similar (Appendix B-6, B-7, B-8, B-12, B-13, B-14, C-6, C-7, C-8, C-12, C-13, C-14). The unusual patterns were the same between the S/MARs as were the patterns involving multiples of Athos and Thalos. Differences occur when analyzing the individual Stowaways. In sequences identified using pTMX4, Hades tended to occur as a singlet whereas sequences identified using pSR14 Hades was found as a doublet approximately half the time. Oleus had more occurrences in pSR14, however there were more occurrences where Oleus was paired in pTMX4. Polyphemus was found more often as a single in pSR14 than pTMX4 where Polyphemus was often in a multiple. In both S/MARs, Polyphemus was usually only a distance of 1 or 2 Stowaways away from another Polyphemus. There was 1 additional sequence containing 1 Stowaway in the sequences identified with pSR14, which the Stowaway was Thalos.

CHAPTER 4

DISCUSSION & CONCLUSIONS

Discussion

The actual sequences of pTMX4 and pSR14 were found in different genetic regions of the genome. The sequence for pTMX4 was isolated near a microsatellite, making this S/MAR part of a highly repetitive genomic region (Christoffers 1998). The sequence for pSR14 was isolated from an amylase gene, α*Amy*3 (Christoffers 1998). Repeat regions tend to have more variation due to the increased amount of mutations, which may explain the large decrease in results of the BLASTn search of pTMX4 compared to pSR14. Since pSR14 is within a gene, there is a lesser likelihood that mutations occur and persist through evolution. The actual sequence for pSR14 is shorter than the sequence for pTMX4. It is easier to find matches in a genome for shorter sequences, so this may also explain the increased quantity of results from the BLASTn search for pSR14.

The BLASTn search resulted in a higher number of initial sequences than the MEGABlast search using pTMX4, however, after the sequences were reduced to sequences containing only fully annotated Stowaways, the quantity was nearly equal. It seems that the MEGABlast search using pTMX4 was able to weed out 30 out of 136 sequences that were not usable for this analysis. Of the sequences identified with pTMX4 in the BLASTn search, 46 of 136 (~34%) contained fully annotated Stowaways. Of sequences identified in the MEGABlast search, 43 of 106 (~41%) had fully annotated Stowaways. The MEGABlast search identifies sequences with an increased amount of similarity to the original sequence searched. An increased proportion of sequences identified by the MEGABlast search to have fully annotated Stowaways suggests that Stowaways have a relationship with pTMX4,

in that most MEGABlast sequences contained where only 3 sequences of the 30 lost did not. This did not appear to be a random loss of sequence since it was not proportional to those containing Stowaways versus those that did not, nor was it proportional to the genera or species involved (Appendix A-1). Unlike pTMX4, pSR14 did not obtain a significantly higher proportion of sequences with fully annotated Stowaways using the MEGABlast search as opposed to the BLASTn search. Rather, the MEGABlast search only resulted in 3 sequences with fully annotated Stowaways out of 50 total sequences, 6%, compared to the 50 sequences with fully annotated Stowaways out of 1000 total sequences, 5%. This 1% difference is considered negligible (Appendix A-1). Future analysis of Stowaways using sequences homologous to pTMX4 should utilize the MEGABlast searches to increase the proportion of results which contain Stowaways compared to results that do not.

More species were identified using pSR14. Since pSR14 is part of a gene and pTMX4 includes a portion of a microsatellite, sequences with pTMX4 are going to be more unique to a species since the repetitive microsatellites will vary more among species than the genes. Among all searches, except the MEGABlast search of pSR14, the same five species were identified to have Stowaways within sequences revealed from using pTMX4 and pSR14. These species included: *Aegilops tauschii, Hordeum vulgare, Triticum aestivum, Triticum monococcum,* and *Triticum turgidum.* When pSR14 was used, there were more species which were not plants or cereal grains than with pTMX4. Similar quantities of sequences were identified in species with fully annotated Stowaways between the two S/MARs in the BLASTn search results. This may be explained by the overlap of some sequences. There were an insufficient number of sequences found in the MEGABlast search of pSR14 to do statistical comparisons to pTMX4.

One common species, *Puccinia tritcina*, is a fungus which attacks wheat plants, *Triticum*. This fungus contains sequences homologous to both S/MARs. A question for further research could analyze why homologous sequences have been identified in this pathogen of wheat and how the

homologous sequences became part of the fungal genome. Research, to date, has not shown that S/MARs have any transposable ability. However, if these S/MARs are associated with Stowaways, as the transposons move from one genome to another, this may function as part of the mechanism by which parts of the original wheat genome may become part of the fungal genome.

The general relationship between the Stowaways and S/MARs was the same in both pTMX4 and pSR14. In all sequence sets, the general pattern appears to be that as the S/MAR homology decreases, the quantity and variety of Stowaways also decreases. The homology of a sequence refers to the level of sequence similarity. As the homology decreases, the sequences in question are less and less similar. This difference in similarity between sequences which represent the same genetic element is often due to mutation, insertion, deletion, and/or genome rearrangements. So, the sequences with larger e-values identified by the BLASTn and MEGABlast searches have more sequence variation to the sequence submitted of pTMX4 or pSR14 for this analysis. If the S/MAR is serving as a target sequence for the insertion of Stowaway-like transposable elements, the more mutations which occur in the S/MAR sequence, the less likely it is to be recognized by the transposon. This would be represented by a smaller quantity of Stowaways in S/MAR sequences which have higher e-values, which this analysis did show. A study by Tikhonov, Bennetzen, and Avramova (2000) found Tourist MITEs can bind to the matrix. As previously mentioned, Stowaway MITEs have a similar structure to Tourist MITEs. The ability for transposable elements to bind to the matrix may explain any apparent associations between S/MARs and transposons. This may also explain one potential function of Stowaway transposable elements.

Of the five species identified with the BLASTn search using pTMX4 which contained fully annotated Stowaways, three were species of *Triticum*. The different species each contain a particular genome arrangement. According to Dvorak, Akhunov, Akhunov, Deal, and Luo (2006), *T. monococcum* contains wheat genome A, *T. turgidum* contains wheat genomes A and B, and *T.*

aestivum contains wheat genomes A, B and D. The results using pTMX4 show evidence supporting that as the quantity and variety of genomes increase in the whole species, the quantity and variety of Stowaways with a given sequence also increases. Though the same five species were identified with the BLASTn search of pSR14, the difference between the species was not the same compared to pTMX4. Rather, *T. monococcum* and *T. turgidum* did not have a significant difference between the Stowaway quantity and variety. *T. aestivum* had the same large increase in Stowaway quantity and variety as was seen with pTMX4. Further research can analyze the origins of each wheat genome compared to the other species identified to contain Stowaways within sequences identified by pTMX4 or pSR14.

There was an exponential relationship represented by the number of Stowaways per sequence for both S/MARs. More sequences had fewer Stowaways per sequence whereas only a few sequences had a high quantity of Stowaways. Therefore, if Stowaways serve as a target sequence for the insertion of other Stowaways, then as the number of Stowaways in a given sequence increased, so would the number of target spots for insertion. This would lead to an increased amount of insertion within the sequence. This may explain the exponential relationship because as the number of Stowaways increased, the target areas for insertion increased as well, leading to an increased amount of new Stowaway insertion. Both S/MARs appeared to increase in the number of Stowaways around the same Stowaway level, 12 and 13. Further research could be performed to analyze the length of each sequence identified by the BLAST searches with respect to the number of Stowaways to further confirm these findings.

Many Stowaway-Stowaway relationships were consistent between sequences identified with the S/MARs. Both S/MARs have Athos occurring often in multiples. Within these multiples, Athos is commonly truncated or fragmented. Often after the truncated Athos, another Athos was present. This suggests that Athos inserts within its own 3' region (Figure 3). This would leave a 3'

fragment too small to be identified and a 5' fragment or a truncated sequence, which was observed in the sequences. Athos may serve as a self-target sequence for reinsertion. Within these multiples, there was one complete Athos present upstream to the Athos in which insertions occurred. This completed Athos may serve as a target, or the pair of Athos transposons may serve as a target for insertion.

Similar to the target for Athos, Thalos and Icarus appear to serve as targets for each other. These two Stowaways were often found in a pair in which at least one of the two was fragmented, suggesting the same pattern as Athos (Figure 3). Unlike Athos, two fragments or a 5' section and 3' section were usually observed. This suggests that Icarus and Thalos insert within each other closer to the middle of the sequence, leaving behind large enough fragments to be identified by either the transposons for insertion or BLAST searches identification.

As more Stowaways insert within each other, the initial Stowaway fragments are driven farther apart from one another. If multiple insertions occur within one Stowaway, this may leave several pieces of DNA too small to identify relative to the databases. Without tracing the evolution of the sequences of DNA in which the Stowaways are present, it cannot be known if or when this "loss" of transposons occurs. The length of sequence fragments of Stowaways could assist in following the patterns of insertion. The models in Figure 3 represent patterns that were observed in more than one occurrence. Data suggesting unique patterns of insertion are also present.

It should be noted that several sequences overlapped between the results of pTMX4 and pSR14 (Appendix A-7). Additional S/MARs need to be analyzed for a relationship with Stowaways to further confirm these findings. Some of the overlap may have affected or help explain the similarities between the relationships found with each set of data.

During the data collection for this analysis, it was noticed that the entry of elements within a sequence such as transposable elements is not standardized. Often the families of transposons or

the common names (Athos, Thalos, etc) were missing. The only pieces of the entry which were consistent included the base pair range and whether the sequence was on the complement strand or not. The category or "feature", as described by NCBI, was inconsistent. Transposable elements were categorized as either "mobile elements", "repeat regions", or "misc features". It is unknown what other inconsistencies existed, such as fragmentation type or if the sequence was fragmented at all. Many occasions where a Stowaway was identified, it was listed with an unknown name or as just a generic Stowaway. This lack of standardization currently inhibits research, such as this analysis.

Conclusions

Based on these results, it can be concluded that a relationship between Stowaways and the S/MARs pTMX4 and pSR14 appears to be present. Additional research to be completed includes an analysis of sequences which were identified by the BLASTn and MEGABlast searches by identifying the location of the actual S/MAR sequence with respect to the Stowaway locations within the sequence identified. Other research to be completed includes the analysis of sequence length and the types of genomes present in the different organisms. Additional research can be done to properly annotate Stowaways which have already been identified, which will allow those Stowaways to be analyzed for patterns.

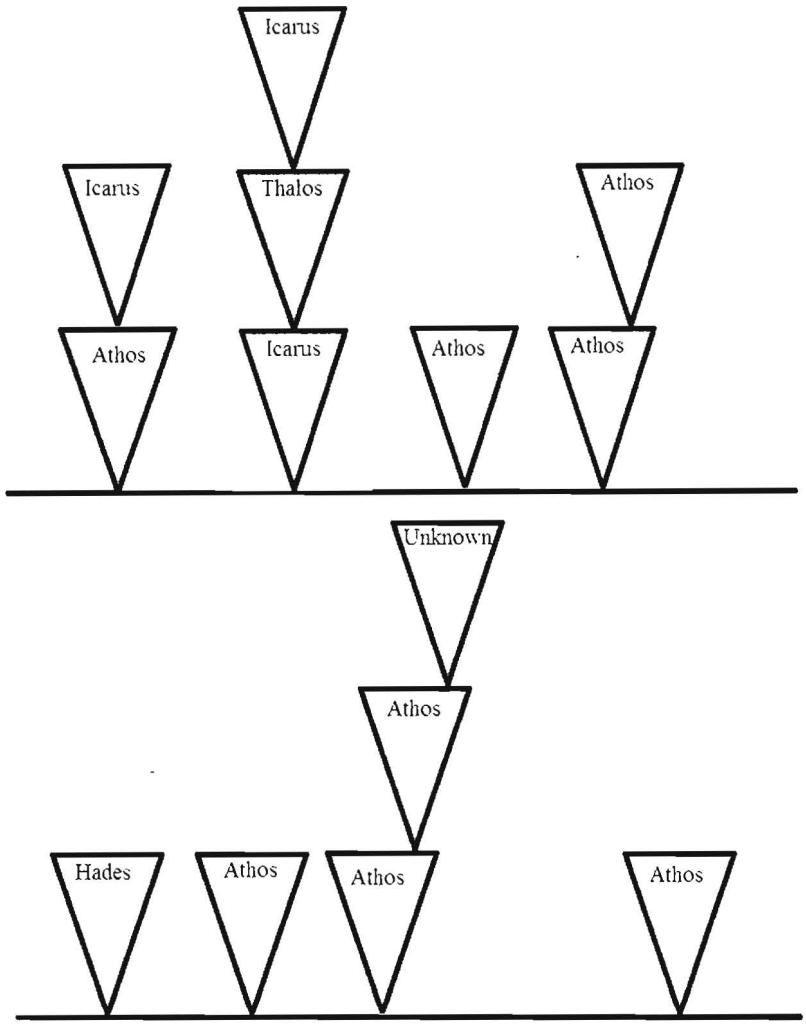


Figure 3. Two models displaying potential insertion patterns for Stowaways within each other. The insertion pattern on the top was predicted using data from the sequence with the accession number CT009625.1, the bottom used data from EU835198.1.

WORKS CITED

Allen, G. C., Spiker, S., &Thompson, W. F. 2000. Use of matrix attachment regions (MARs) to minimize transgene silencing. *Plant Molecular Biology,* **43:** 361-376.

Baulcombe, D.C., Huttly, A.K., Martienssen, R.A., Barker, R.F., & Jarvis, M.G. 1987. A novel wheat α-amylase gene (α-*Amy3*). *Mol. Gen. Genet.*, **209:** 33-40.

Brouwer, C., Bruce, W., Maddock, S., Avramova, Z., & Bowen, B. 2002. Suppression of transgene silencing by matrix attachment regions in maize: a dual role for the maize 5' *ADH1* matrix attachment region. *The Plant Cell,* **14:** 2251-2264.

Bureau, T. E, Ronald, P. C., & Wessler, S. R. 1996. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proc. Natl. Acad. Sci. USA*. **43:** 8524-8529.

Bureau, T. E., & Wessler, S. R. 1994. Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *The Plant Cell,* **6:** 907-916.

Brown, T. A. 2007. "Chapter 18: How Genomes Evolve". *In:* Genomes. Garland Science Publishing, New York, NY. pp. 563-593.

Christoffers, M. J., 1998. Identification and characterization of matrix attachement regions in wheat. Ph.D. dissertation, Dissertation Chair: Gustafson, J. P., USDA-ARS-PGRU & Department of Agronomy, University of Missouri-Columbia, Missouri.

Delongchamp, S. 2007. Bioinformatics analysis of predicted S/MARs and associated Stowaway transposon location in the gramineae. M.Sc. thesis, Thesis Advisor: C.A. Blakey, Department of Biology, Ball State University, Muncie, IN.

Durand, P. M. & Michod, R. E. 2010. Genomics in the light of evolutionary transitions. *Evolution.* **64(6):** 1533-1540.

Dvorak, J., Akhunov, E. D., Akhunov, A. R., Deal, K. R., & Luo, M-C. 2006. Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. *Mol. Biol. Evol,* **23(7):** 1389-1396.

Feschotte, C., Osterlund, M. T., Peeler, R., & Wessler, S. R. 2005. DNA-binding specificity of rice *mariner*-like transposases and interactions with *Stowaway* MITEs. *Nucleic Acids Research,* **33(7):** 2153-2165.

Feschotte, C. Swamy, L. & Wessler, S. R. 2002. Genome-wide analysis of *mariner*-like transposable elements in rice reveals complex relationships with *stowaway* miniature inverted repear transposable elements (MITEs). *Genetics,* **163:** 747-758.

Gasser, S. M. & Laemmli, U. K. 1986. The organization of chromatin loops: characterization of scaffold attachment site. *The EMBO Journal,* **5(3):** 511-518.

Heng, H. Q., Goetze, S., Ye, C. J., Liu, G., Stevens, J. B., Bremer, S. W., Wykes, S. M., Bode, J., & Krawetz, S. A. 2004. Chromatin loops are selectively anchored using scaffold/matrix-attachment regions. *Journal of Cell Science,* **117:** 999-1008.

Lewin, B. 2008. "Chapter 21: Transposons". *In:* Genes IX. Jones and Bartlett Publishers, Inc., Boston, MA. pp. 521-549.

Lewin, B. 2008. "Chapter 22: Retroviruses and Retrotransposons". *In:* Genes IX. Jones and Bartlett Publishers, Inc., Boston, MA. pp. 550-569.

Liebich, I., Bode, J., Frisch, M., & Wingender, E. 2002. S/MARt DB: a database on scaffold/matrix attachment regions. *Nucleic Acids Research,* **30(1):** 372-372.

Liebich, I., Bode, J., Reuter, I., & Wingender, E. 2001. Evaluation of sequence motifs found in scaffold/matrix-attachment regions (S/MARs). *Nucleic Acids Research,* **30(15):** 3433-3442.

Linnemann, A. K., Platts, A. E., & Krawetz, S. A. 2009. Differential nuclear scaffold/matrix attachment marks expressed genes. *Human Molecular Genetics,* **18(4):** 645-654.

Macas, J., Koblizkova, A., & Neumann, P. 2005. Characterization of *Stowaway* MITEs in pea (*Pisum sativum* L.) and identification of their potential master elements. *Genome.* **48:** 831-839.

Rudd, S., Frisch, M., Grote, K., Meyers, B. C., Mayer, K., & Werner, T. 2004. Genome-wide in silico mapping of scaffold/matrix attachment regions in arabidopsis suggests correlation of intragenic scaffold/matrix attachment regions with gene expression. *Plant Physiology,* **135:** 715-722.

Tikhonov, A. P., Bennetzen, J. L., & Avramova, Z. V. 2000. Structural domains and matrix attachment regions along collinear chromosomal segments of maize and sorghum. *The Plant Cell,* **12:** 249-264.

Turcotte, K., Srinivasan, S., & Bureau, T. 2001. Survey of transposable elements from rice genomic sequences. *The Plant Journal,* **25(2):** 169-179.

BIBLIOGRAPHY

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* **25(17):** 3389-3402.

Chang, R. Y., O'Donoughue, L. S., & Bureau, T. E. 2001. Inter-MITE polymorphisms (IMP): a high throughput transposon-based genome mapping and fingerprinting approach. *Theor Appl Genet,* **102:** 773-781.

Drunen, C. M. v., Oosterling, R. W., Keultjes, G. M., Weisbeek, P. J., Driel, R. v., & Smeekens, S. C. M. 1997. Analysis of the chromatin domain organization around the plastocyanin gene reveals an MAR-specific sequence element in *Arabidopsis thaliana*. *Nucleic Acids Research,* **25(19):** 3904-3911.

Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics,***9:** 397-405.

Feschotte, C. & Pritham, E. J. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics,* **41:** 331-368.

Feschotte, C. & Wessler, S. R. 2002. Mariner-like transposases are widespread and diverse in flowering plants. *Proc. Natl. Acad. Sci,* **99(1):** 280-285.

Frisch, M., Frech, K., Klingenhoff, A., Catharius, K., Liebich, I., & Werner, T. 2001. In silico prediction of scaffold/matrix attachment regions in large genomic sequences. *Genome Research,* **12:** 349-354.

Goodrich, J. & Tweedie, S. 2002. Remembrance of things past: chromatin remodeling in plant development. *Annu. Rev. Cell Dev. Biol.,* **18:** 707-746.

Han, Y., Burnette III, J. M., & Wessler, S. R. 2009. TARGeT: a web-based pipeline for retrieving and characterizing gene and transposable element families from genomic sequences. *Nucleic Acids Research*, **37(11):** 1-13.

Hubner, M. R. & Spector, D. L. 2010. Chromatin Dynamics. *Annu. Rev. Biophys*, **39:** 471-89.

Jiang, N., Feschotte, C., Zhang, X., & Wessler, S. R. 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Plant Biology*, **7:** 115-119.

Kanazawa, A., Akimoto, M., Morishima, H., & Shimamoto, Y. 2000. Inter- and intra-specific distribution of Stowaway transposable elements in AA-genome species of wild rice. *Theor Appl Genet,* **101:** 327-335.

Kidwell, M. G., & Lisch, D. 1997. Transposable elements as sources of variation in animals and plants. *Proc. Natl. Acad. Sci.,* **94:** 7704-7711.

Mao, L., Wood, T. C., Yu, Y., Budiman, M. A., Tomkins, J., Woo, S., Sasinowski, M., Presting, G., Frisch, D., Goff, S., Dean, R. A., & Wing, R. A. 2000. Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. *Genome Research,* **10:** 982-990.

Polidoros, A. N. & Scandalios, J. G. 1998. Circadian expression of the maize catalase *Cat3* gene is highly conserved among diverse maize genotypes with structurally different promoters. *Genetics*. **149:** 405-415.

Takahashi, H., Akagi, K., Mori, K., Sato, K., and Takeda, K. 2006. Genomic distribution of MITEs in barley determined by MITE-AFLP mapping. *Genome,* **49:** 1616-1620.

DETAILED LIST OF APPENDICES